

ANALYSIS OF PORTABLE SYSTEM FOR SOUND ACQUISITION OF VEHICLES POWERED BY INTERNAL COMBUSTION ENGINES

Marko Milivojčević¹, Emilija Kisić², Dejan Ćirić³

¹Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, Belgrade, Serbia

²Metropolitan University, Faculty of Information Technology, Belgrade, Serbia

³University of Niš, Faculty of Electronic Engineering in Niš, Niš, Serbia

Abstract. *In this paper a portable system for acquisition of sound generated by passenger vehicles powered by internal combustion engines is described and analyzed. The acquisition system is developed from scratch and tested in order to satisfy the requirements such as high-quality of audio recordings, high mobility, robustness and privacy respect. With this acquisition system and adequate signal processing, the main goal was to collect a large amount of clear audio recordings that will form a quality dataset. In further research, this dataset will be used for machine learning model training and testing, i.e. for developing a system for automatic recognition of the type of car engine based on fuel.*

Key words: *acoustic based acquisition system, dataset, audio signals, internal combustion engines*

1. INTRODUCTION

Applications of artificial intelligence algorithms to audio signals are becoming more numerous over time [1-4]. Sound classification, audio event detection and audio scene recognition are examples of tasks that are successfully realized in practice by applying machine or deep learning [5, 6]. In this context, machine and deep learning could be used to identify the type of internal combustion engine with regard to the fuel based on the sound generated by the engine. Namely, the sound of these engines differs depending on the used fuel - petrol (gasoline) or diesel. Human ear can recognize this sound difference, that is, whether it is a petrol or diesel engine's sound. Those facts and the need to classify passenger vehicles by fuel as a result of improved environmental standards [7] have served as major pillars of the present research. Its main aim is to develop a system for automatic recognition of engine type based on sound generated by the engine, that is, to

Received November 09, 2022; revised January 18, 2023; accepted February 06, 2023

Corresponding author: Marko Milivojčević

Academy of Technical and Art Applied Studies, School of Electrical and Computer Engineering, Belgrade, Serbia

E-mail: markom@viser.edu.rs

© 2023 by University of Niš, Serbia | Creative Commons License: CC BY-NC-ND

build a machine/deep learning model that will be able to recognize the type of engine with high accuracy, where an input to the model will be the engine sound.

Since the successful implementation of machine/deep learning requires an adequate dataset (containing, in this case, audio samples), a specialized acquisition system has been developed for this purpose. Details of the development of such an acquisition system for the collection of audio samples of the passenger vehicles powered by internal combustion engines are presented here.

The first requirement that the acquisition system should satisfy is the automation, because manual collection of a large number of samples would require a lot of time and might introduce certain differences in conditions during the acquisition. Then, the collected data should meet the requirements for quality, duration and invariability of environmental conditions in order to provide the reliable information regarding the acoustic characteristics.

The paper is divided into several sections. The technical characteristics of the system, hardware configuration and selection of components as well as acquisition procedure and processing of the collected audio signals are presented in the section related to methodology. The section describing the results provides a tabular presentation of the system efficiency for three cases of time interval between the start of detection of two consecutive vehicles, as well as the presentation of audio signals in the time and spectral domain as a measure of validity of the obtained images for further analysis with a machine or deep learning system. The paper ends with concluding remarks.

2. ACQUISITION SYSTEM AND PROCEDURE DESCRIPTION

In the earlier phases of the research, the influence of microphone position in the area below the engine compartment on the characteristics of audio recordings was analyzed in detail [8]. As a result, it was determined that the basic characteristics of the audio signal varied minimally independently where the measuring microphone was placed as long as the microphone was directly below the engine compartment [8]. In that regard, depending on a vehicle, the target area where the microphone can be placed below the engine is approximately 1.2 m by 1.2 m. Because of that, it is possible to collect relevant audio samples regardless of the exact position of the vehicle when it is stopped above the microphone. Based on the previous findings, the acquisition system uses a microphone positioned in the area below the engine compartment chosen as the most suitable area in terms of "purity" of sound [8, 9], and audio recording begins only after the presence of the vehicle is detected. In this way, audio samples of engine operation in the idle mode are collected, without the microphone itself being positioned on the vehicle.

The system has been developed to be mobile, so that it can be set up independently of availability of power sources, and it is fully designed to run on battery power. Additionally, the system is designed to be autonomous, i.e., not to require human presence during operation. As the system has limited memory space, it was necessary to develop several verification steps before the current audio sample was written in the memory. Specifically, this system has four levels of verification before storing the audio recording, which resulted in a dataset of recordings that contains only sounds of interest, i.e., engine operation. When the system is applied in real conditions involving presence of interfering sources of noise and different engine load modes, despite a large number of successfully collected audio recordings, some recordings containing not only the desired engine mode but also other engine modes appeared

in the formed dataset. So, it was necessary to develop a procedure that detects and then extracts the idling mode of the internal combustion engine. In order to have as much autonomy of the system as possible, the requirement for minimum energy consumption conditioned the application of the simplest possible procedure for separating the desired engine mode. Thus, the procedure of extracting the engine idle mode applied here is based on the audio signal processing in the time domain, i.e., usage of signal envelope. It is worth mentioning that the number of recordings containing only the engine idling mode is also affected by the minimum time period between the start of detection of two consecutive vehicles.

2.1. Acquisition system

The main goal of collecting audio samples of engine operation is to make a dataset containing sounds of passenger vehicles recorded in real conditions. In this way, the future classification system will be able to properly work in such conditions, as those at entrances to underground garages, toll plazas, gas stations, etc. The generated dataset of audio samples should preferably have such characteristics that will enable its usage in different machine and deep learning approaches [10]. They include support vector machine (SVM) [11], k-nearest neighbors (k-NN) [12], deep forest [13] or various deep neural network architectures as multi-layer perceptron [14] or convolutional neural network [15]. For this purpose, the audio samples may be transformed either into selected set of features or images, such as spectrogram-like images, or they may be used in the existing format (raw audio signals).

The entrance to the underground garage with a ramp was chosen as the most suitable space for collecting audio samples, where it is necessary to stop the vehicle until the driver takes the card / token. During this period, the car is static and idling. Even if it has a start / stop system, it will run in idle mode for a certain period of time. In addition, in such a situation, the movement of the vehicle is so directed that there is no possibility of mechanical damage to the microphone and sensor that are placed on the ground in the space between the wheels. The block diagram of the system is presented in Fig. 1, and the realized system in a laboratory environment is shown in Fig. 2.

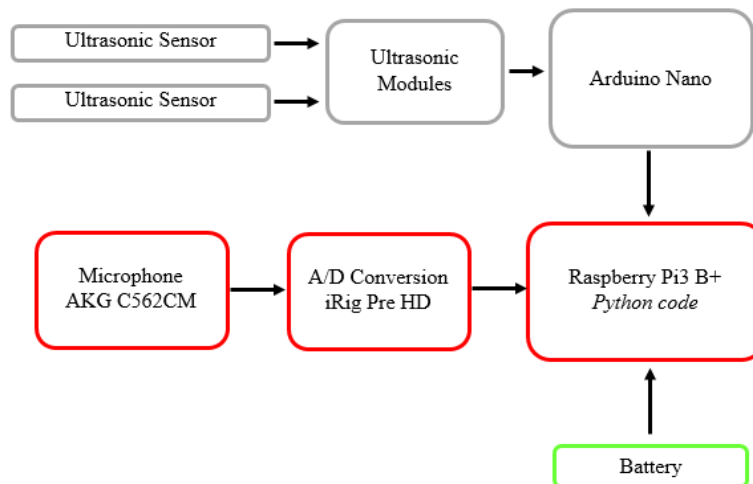


Fig. 1 Block diagram of the sound acquisition system

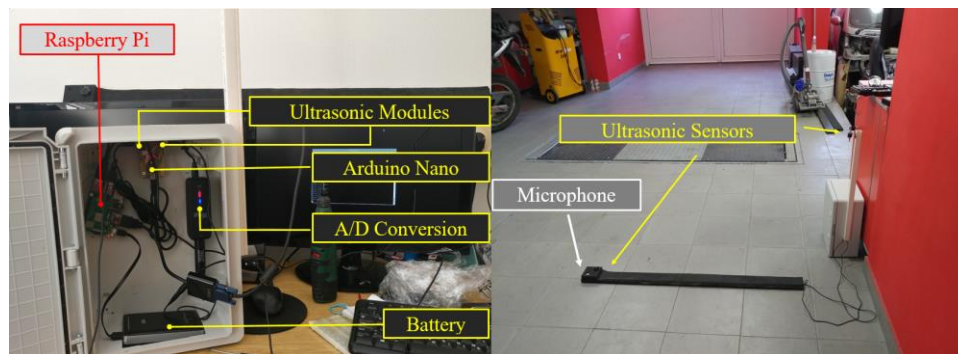


Fig. 2 Realized acquisition system in laboratory environment

The system has been developed so that the presence of vehicles is detected with the ultrasonic sensors before the process of recording the engine operation sound begins (the first level of verification). Ultrasonic sensors are primarily selected as sensors that, unlike widely used cameras, do not affect user privacy. Also, these sensors that are among the cheapest sensors on the market have very low power consumption, and they are accurate enough to detect vehicles. This type of vehicle detection enables the installation of the system almost everywhere because there is no possibility of interference with any existing induction sensors at the entrance ramp and violations of the law related to user privacy. In order to avoid detection of objects that are not vehicles of interest, two sensors are used. The sensors are positioned so that one measures the distance along the horizontal (x) axis and the other one along the vertical (y) axis. The plane formed by the ultrasound sensors is not perpendicular to the direction of vehicle movement, as shown in Fig. 3.

By using the sensors placed in the described way, the possibility of detecting two-wheelers and pedestrians that might also show at the ramp is eliminated. Namely, due to the position and orientation of the sensor placed on the ground, two-wheelers can only be detected if they pass directly above, i.e., over the sensor. However, even in such a case, they will not meet the distance requirement from the side sensor, if they move in the intended direction of entering the garage. If the sensors were located in the same plane, then it would theoretically be possible for a motorcycle to be oriented perpendicularly in reference to the intended direction of movement of the vehicle, i.e., above the ground sensor and facing the side sensor with the front or rear wheel. By positioning the sensors in two planes, a motorcycle would have to be in an almost impossible position to enter the garage, i.e., it would need to hit the ramp in order to satisfy the condition of the vehicle presence on both sensors. In a similar manner, a pedestrian who is above the ground sensor could move in tandem with another pedestrian who would satisfy the condition of the side sensor if both sensors were in the same plane. However, if the distances are measured in different planes, it would be more difficult and less likely to meet the condition of the vehicle presence on both sensors.

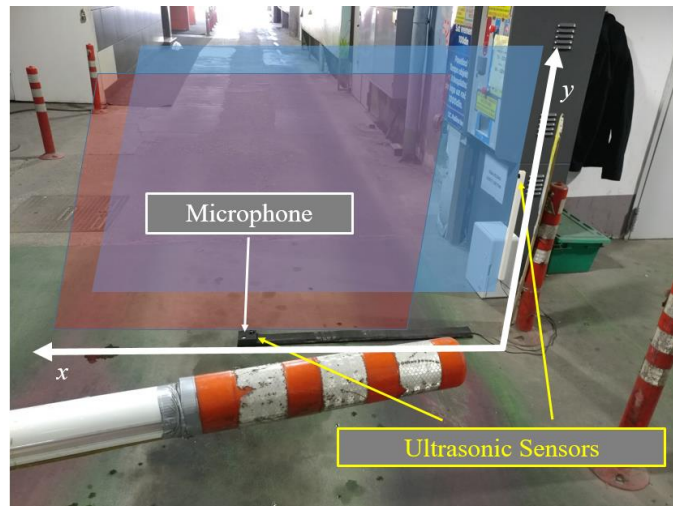


Fig. 3 The acquisition system positioned at the entrance to the underground garage, where horizontal (x) and vertical (y) axis as well as horizontal and vertical plane, which is also the plane formed by the axes, are presented

Readily available waterproof ultrasonic distance measurement modules containing an ultrasonic sensor JSN-SR04T, whose specification is given in [16], are used in the acquisition system. These modules, that is, sensors are controlled by a microcontroller within the Arduino Nano platform [17], where distances are set for the specific measurement case. Distance measurement is realized by the short-term emission of an ultrasonic signal triggered by Arduino, after which Arduino measures the time until the reflected signal appears. The distance to an obstacle is calculated based on the measured time required for the signal to reach the obstacle and then return, and based on the speed of sound in the air. Since measurement of the distance to the vehicle does not require precision greater than 1 cm, the best results were obtained by a trigger signal lasting 10 microseconds. Thus, if both sensors detect an object (the horizontal sensor at distance less than 80 cm and the vertical sensor at distance less than 40 cm), the microcontroller registers a vehicle presence and sends this information using serial communication to the Raspberry Pi computer [18]. This computer represents the central part and heart of the acquisition system. The reason for using an additional microcontroller in addition to the Raspberry Pi, which can also control and read ultrasonic sensors, is the need to detect vehicle presence continuously, i.e., in parallel with recording the audio. By having both the Raspberry Pi and the additional microcontroller (Arduino Nano), two activities – vehicle detection and audio recording, supposed to be done in parallel, can be realized in an easier and more reliable way.

Each audio sample is recorded with an omnidirectional microphone that is placed on the ground in the area below the vehicle. In this way, in almost all cases, the microphone is positioned directly below the vehicle's engine after the vehicle is stopped in front of the ramp. In order to obtain the highest quality audio recordings, the AKG C562CM omnidirectional microphone is used, with the specifications that are listed in [19] and presented in Fig. 4.

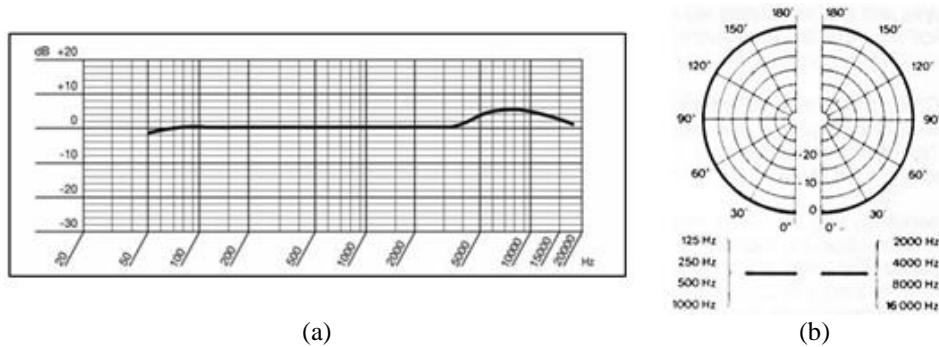


Fig. 4 Characteristics of AKG C562CM microphone: (a) frequency and (b) polar response [19]

The microphone and ultrasonic sensor that measures the vertical distance are placed in a purpose-made cable protector (Fig. 5a) made of industrial rubber with a hardness of 90 Shora. For the purpose of collecting audio recordings, the edges of this cable protector had to be processed at an appropriate angle so that the sound of wheel crossing over it should be negligible in the recordings. The processing angle was determined empirically and was approximately 150° . In addition to protecting the cables that connect the microphone and ultrasonic sensor to the rest of the system, the cable protector is designed to protect both the microphone and sensor in the case that the vehicle wheel passes directly over them (Fig. 5b). When the cable protector is placed at the measuring position, it is not necessary to fasten this guide to the base, because it is not subject to slipping and moving due to the structure of the rubber and its width of 20 cm. The guide was at almost the same position during the acquisition independently on how large and heavy were the vehicles passing over.

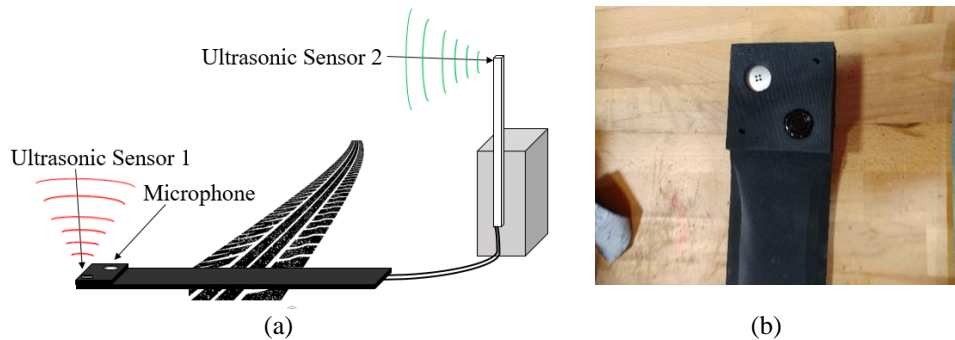


Fig. 5 Cable and ultrasonic sensor protection: (a) purpose-made cable protector and (b) microphone/sensor protection

The hardware limitations of the Raspberry Pi computer in terms of maximum sampling frequency and number of bits for audio signal quantization as well as the need for microphone phantom power resulted in the insertion of an A/D converter between the microphone and the Raspberry Pi computer. For the purpose of A/D conversion and microphone power supply, a dedicated high-quality audio interface iRig Pre HD is employed, which is also a battery-powered device whose specifications are given in [20]. Additionally, the use of an external

A/D converter enables the Raspberry Pi to run at lower processing power and lower power consumption. On the Raspberry Pi computer, the developed Python code is run after the power is turned on. Within this code, the serial communication via USB port is listened to in order to receive the information from the Arduino about the vehicle presence. When the vehicle is detected, a series of processes are realized that are described in the next subsection.

2.2. Acquisition procedure

After the vehicle presence is detected, and in order to save the battery, the Raspberry Pi starts the microphone listening mode via the A/D interface. Only when the detected sound level is above the set threshold, the storage of the stream in the buffer will begin (the second level of verification). The threshold level is determined empirically at 74 dB. In this way, an accidental excitation of the sensors that can be caused by the passing of a pedestrian, dog or cat is avoided. The audio recording duration is initially set to 5 s and after the time has elapsed, the stream stops. In order to avoid an accidental excitation potentially caused by the passing of a motorcycle, the stream is additionally checked after stopping it. Namely, at the location where the samples were taken, and in most of the underground garages, motorcycles are allowed to enter without any obstruction next to the ramp, so they are not stopped at the entrance. The mentioned check is performed simply - after two seconds from the beginning of the stream the signal level is checked whether it is above the threshold set in the previous step or not (the third level of verification). If the threshold condition is met in that segment of the stream, it is stored as a wav file on the SD card. The entire procedure is shown as the flowchart given in Fig. 6.

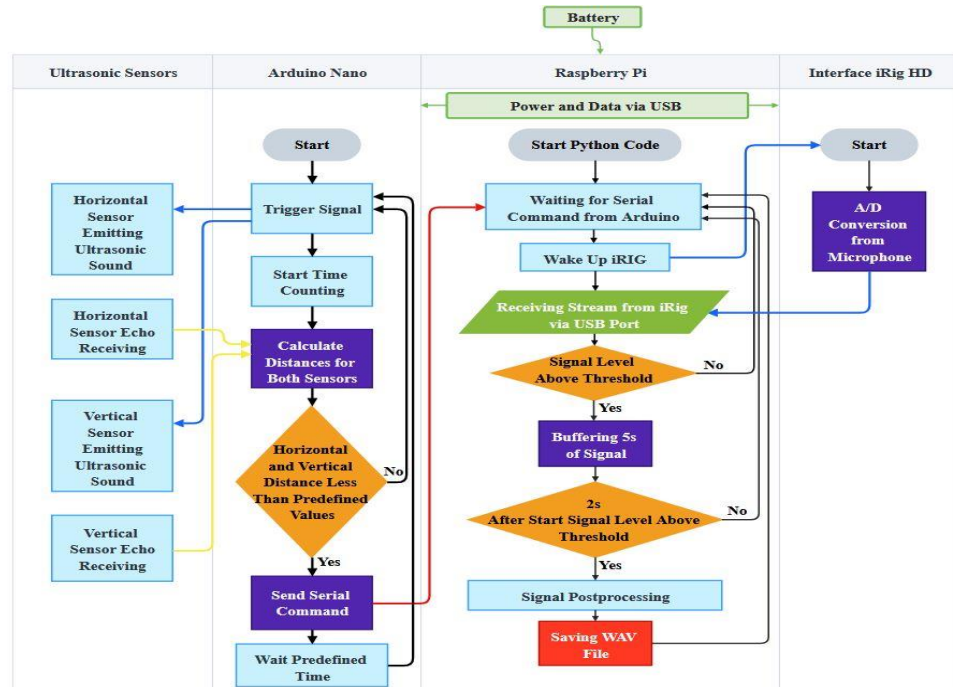


Fig. 6 Acquisition procedure flowchart

The fourth level of verification is a specially developed algorithm where only the engine idle mode (stationary signal) is extracted from the existing wav file. The description of this procedure is given in the next subsection.

The initial installation of the system at the entrance ramp of the underground garage showed that the system detected only vehicles and that the audio recordings contained only signals originating from internal combustion engines. However, the waiting time of vehicles above the microphone varied considerably from case to case. Due to this phenomenon, three different approaches for audio signal recording (A, B and C) were applied based on the activities of ultrasound sensors. In the first one (A), it was defined that after detecting the object (vehicle), the ultrasound sensors remained inactive for 5 s until the rest of the system finished the audio sample recording. In the approach B, a fixed time of 5 s of sensor inactivity after detection was replaced by the time of 8 s. The third approach (C) is related to the situation where the sensors were constantly active in order to detect when the vehicle left the space above the microphone, thus not sending a command to the rest of the system to start the next recording. If the sensors in two successive iterations separated in time for 50 microseconds detected the absence of a vehicle, the system interpreted this situation as the vehicle had left the position. This is important because occasionally one of the sensors measures greater distance to a vehicle caused by the higher-order reflections of ultrasonic waves, due to the long waiting of the vehicle. This is interpreted as non-compliance with the presence condition. Such a phenomenon is attributed to the dispersion of ultrasonic waves that can occur due to the shape of the vehicle's body. During the system testing, it was shown that this phenomenon was rare.

In terms of the negative effects of constant exposure to the ultrasonic waves, the used ultrasonic sensors are of very low power, designed to measure distances of up to 4.5 m, which means that the signal level can be negligible at longer distances due dispersion. If we look at the configuration of the entrances to the underground garages, the width of the passage for vehicles must be at least 3 m. In this way, if a pedestrian passage exists, it can only be found at a distance greater than 3 m from the sensor. Besides, within the few hours of the acquisition, fewer than 10 passengers were seen in the pedestrian passage, but being further than 5 m from the sensors.

2.3. Extraction of idling mode of operation

In order to extract the stationary part of each recorded audio signal that corresponds to the engine idle by the signal processing in the time domain applied here, it is necessary to determine the threshold (time moment) after which the non-stationary part of the signal should be rejected. Due to the nature of the problem, the stationary part of the signal always appears at the signal beginning, see Figs. 7, 8 and 9 given in the next section. There are no cases where the idling occurs later (in the middle or at the end of the signal). So, it is clear that the threshold needs to be found at a certain time point after the signal starts, i.e., at the first moment when the signal becomes non-stationary.

Based on the analysis of the waveforms of the recorded signals in the time domain, it is noticed that at the moment when the signal ceases to be stationary, its amplitude abruptly increases. Thus, at that moment, there is a noticeable increase (jump) in the signal envelope. The idea for extracting a stationary part of a recorded signal is based on generating the signal envelope and calculating the difference between the current and previous envelope values along the envelope. While the signal is stationary, the difference between the current and

previous envelope values is expected to be small. On the other hand, at the moment when the signal ceases to be stationary, the difference between the current and previous value of the envelope must be significantly greater than the difference at time instants before that moment. The first time instant from the beginning to the end of the signal where there is a significant increase in the difference between the current and the previous envelope value is a candidate for setting the threshold. This significant increase needs to be quantified.

If the signal envelope is denoted as $Env(t)$ and the threshold representing the upper time limit of the stationary signal part as t_L , the threshold itself can be determined as:

$$t_L = \min \{ Env(t) - Env(t-1) > A \} \cdot \left(\frac{t_s}{N_f} \right) \quad (1)$$

where t_s denotes the duration of the signal, and N_f is the number of frames in which the signal maxima are calculated in the procedure of generating the signal envelope. A is a constant having the value of 0.1 determined empirically. Since it is necessary to set the threshold at the first time instant after the envelope jump looking from the signal beginning to its end, the smallest value that satisfies the condition in (1) is taken as the threshold t_L . More precisely, since the time variable t is given in frames used for generating the signal envelope, the condition $\min \{ Env(t) - Env(t-1) > A \}$ returns an envelope frame in which there is an envelope jump indicating a transition from stationary to non-stationary part of the signal. In order to obtain the exact time instant for setting the threshold, it is necessary to normalize the obtained envelope frame value by t_s/N_f . In our case, the frame size for generating the envelope is 4000 samples with the frame overlap of 1000 samples. This means that the resolution for setting the threshold t_L is determined by the frame size, which can be chosen in accordance with the nature of the analyzed signal.

3. ANALYSIS OF RECORDED AUDIO SIGNALS

Positioning the acquisition system at the entrance of the underground garage with a ramp where it is necessary to stop a vehicle in order to take a token gave the results above the expectations in terms of the quality and number of audio recordings. These recordings have the following parameters: sampling rate of 44.1 kHz, the bit depth of 16 bits, fixed duration of 5 s resulting in a file size of approximately 431 kB, which provides the possibility of storing approximately 67800 audio samples assuming the effective storage space of 28 GB on the 32 GB SD card. The power supply used a power bank with a capacity of 10 Ah, consumed about 20% of the capacity for 2 hours of recording, showing that the system is able to function with this power supply for about 10 hours in a completely autonomous way.

In parallel with the autonomous operation of the system, manual records of the engine type by fuel were made, meaning that the samples were labeled manually. This was done to identify the possible error, e.g. audio recording that would be unusable due to excessive noise of the environment that might be present indoors typically coming from the garage ventilation. This case did not happen in practice as a result of a correctly set threshold that determines the beginning of the recording. Considering the three approaches mentioned above (A, B and C), after analyzing the recordings, the most important fact is that no vehicle passed by the acquisition system without triggering the system to record the sound of its engine operation. Also, events other than passenger vehicles passing by did not falsely trigger the system, and a completely blank recording was not obtained. Table 1 provides a comparative overview of

these three approaches in terms of the number of samples collected as well as the usability of the samples. It is worth mentioning, that during the collection of audio samples, a very small percentage of vehicles belonged to the older generation of vehicles. The majority of diesel vehicles belonged to the generation of common rail type injection, while the majority of gasoline vehicles had multipoint indirect injection.

Table 1 gives the total number of audio recordings and number of useful audio recordings. Here, the latter contain the engine idling mode sounds, while the rest of recordings still contain the vehicle engine sounds, but not the idling mode of operation – instead they contain the sound of a vehicle leaving the ramp. Large majority of recordings are the useful recordings, and its percentage in reference to the total number of recordings is above 90%, where this percentage is the highest for the ultrasonic detection approach C, and it is close to 97%.

By comparing three ultrasonic detection approaches from Table 1, it can be noticed that the approach A with a fixed time interval of detection (sensor inactivity) of 5 s gave the most audio samples, as many as 202% of useful recordings in relation to the number of vehicles. This approach is primarily suitable for generating the largest possible dataset, but it is not suitable from the point of view of efficient usage of the storage resources. If the system is used employing this approach for detection and recognition of the engine type in real conditions, there will be cases where the same vehicle is detected more than once. Strictly speaking, this increased number of recorded audio signals for some vehicles could have certain detrimental effects on the machine/deep learning due to over-representation of these vehicles in comparison to others. Although the number of recordings for majority of vehicles is up to two, these effects will be analyzed in the next phases of the research. Besides, if necessary, the redundant recordings for the same vehicle could easily be removed from the dataset according to the time of recording.

Table 1 Comparative overview of three different detection approaches (A, B and C) in terms of the number of samples collected as well as the usability of the samples

	A (sensors inactive for 5 s after vehicle detection)	B (sensors inactive for 8 s after vehicle detection)	C (continuous detection of vehicles by sensors)
Number of vehicles that passed through the acquisition system	50	100	100
Number of detected vehicles	50	100	100
Total number of audio recordings	111	143	122
Number of useful audio recordings	101	133	118
Number of idle mode records only (without any additional processing)	69	97	109
Percentage of vehicles detected	100%	100%	100%
Percentage of useful recordings in relation to the total number of recordings	90.99%	93%	96.72%
Percentage of useful recordings in relation to the number of sampled vehicles	202%	133%	118%
Percentage of recordings of idle mode only without additional processing in relation to the number of sampled vehicles	138%	97%	109%
Percentage of recordings not requiring the fourth level of verification in relation to the total number of recordings	62.16%	67.83%	89.34%

The approach B (time interval of sensor inactivity of 8 s) also gave good results in terms of the number of vehicles detected and the amount of audio recordings. However, it has the lowest percentage of recordings of idling mode only without additional processing compared to the number of sampled vehicles. This approach has more efficient usage of memory resources compared to the first approach.

The most complex approach (C), continuous detection with the recognition of the next vehicle, gave the least audio recordings in relation to the number of detected vehicles. On the other hand, this approach is the most efficient in terms of memory utilization, achieving a high percentage of clean recordings. In this way, the lowest redundancy among samples and the highest percentage of useful recordings in relation to the total number of recordings were obtained. The latter led to the least need for additional processing (saving CPU resources) and additional power from the power supply.

Within all three approaches from A to C, one or two audio recordings per vehicle were obtained for the majority of vehicles. Here, the first recording represented the engine idling stationary mode without exceptions, Fig. 7. In most of the samples, the second recording (in some cases the last one) partially contained the engine idling mode followed by an increase in the crankshaft speed and partial engine load mode in order to accelerate the vehicle, as shown in Figs. 8 and 9. There were no cases where the partial load mode of the engine appeared before the idling mode in the recordings. In these three figures (Figs. 7, 8 and 9), the audio signals of approximately the same generation of vehicles are presented. Here, the signals' amplitudes are normalized; hence the focus is on differences in the signal level caused by the change in operating mode.

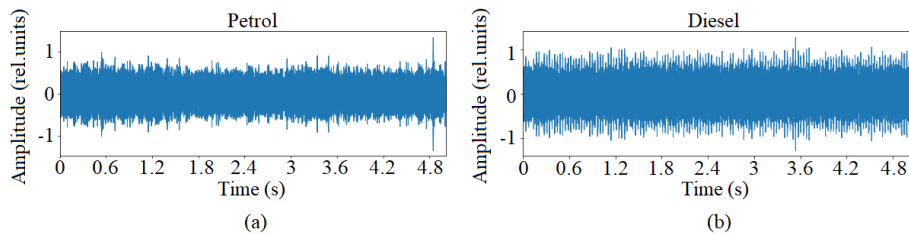


Fig. 7 Audio signal of (a) petrol and (b) diesel engine at idle, without changing the mode

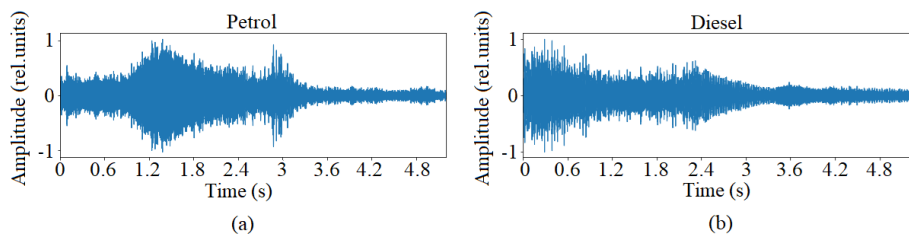


Fig. 8 Audio signal of (a) petrol and (b) diesel engine having early operation mode change from idle to load mode (during the recording interval)

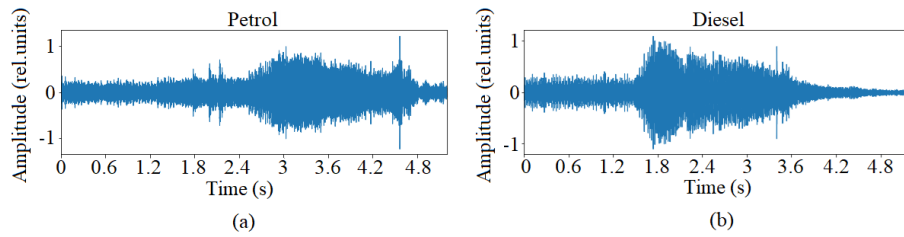


Fig. 9 Audio signal of (a) petrol and (b) diesel engine having late operation mode change from idle to load mode (during the recording interval)

Calculation of the threshold (i.e., the time instant of the audio signal until which the engine is in the idling mode) used for extraction of idling mode of operation is illustrated in Figs. 10 and 11, where the threshold is marked with a purple vertical line. The terms “early” and “late” are related to the cases where the operation mode change happens earlier (up to 1 s) and later (after 1 s) in the recorded signal, respectively. In the recorded signals where there is no change in the engine operation mode, the threshold (cutoff time) could not be determined in the described way. In such a case, the entire audio track is selected as an engine idle, and is used for further analysis and processing.

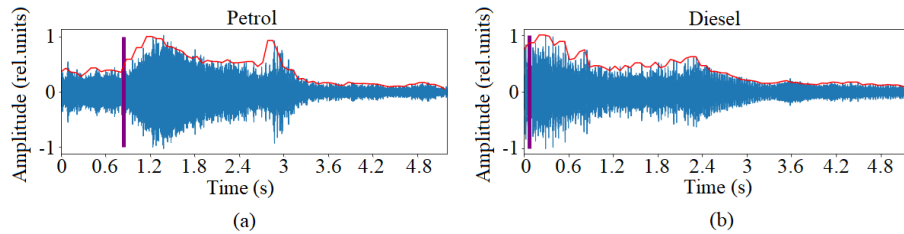


Fig. 10 Waveform and envelope of the audio signal of (a) petrol and (b) diesel engine with an early change of operation mode (the threshold is marked by a vertical line)

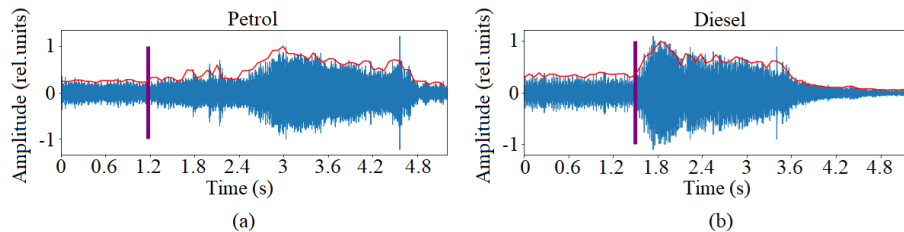


Fig. 11 Waveform and envelope of the audio signal of (a) petrol and (b) diesel engine with a late change of operation mode (the threshold is marked by a vertical line)

The waveforms of the characteristic audio signals extracted in the described way are presented in Figs. 12 and 13. For the presented case of the vehicle using diesel fuel where an early operation mode change (almost at the very beginning of the recording) occurred, the calculated threshold (cutoff) time was also very close to the beginning of the signal

(Fig. 10b), which means that this recording is rejected using the function for checking the duration of the stationary mode. This duration can be set according to the requirement related to the minimal length of the signals. Depending on a particular need, the signal length may be either shorter or longer. In the present case, the duration of the stationary mode is set to 0.5 s meaning that the minimal signal length is 0.5 s.

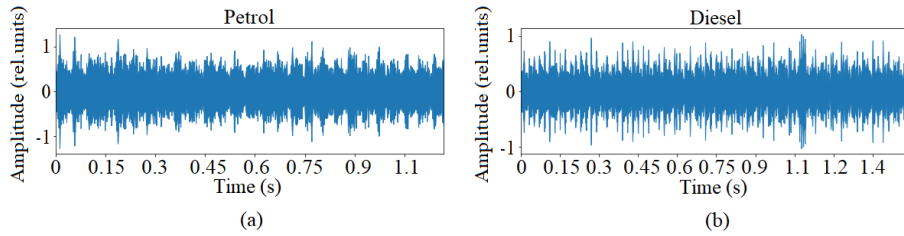


Fig. 12 Audio signal of (a) petrol and (b) diesel engine at idle extracted from the recordings with a late change of operation mode

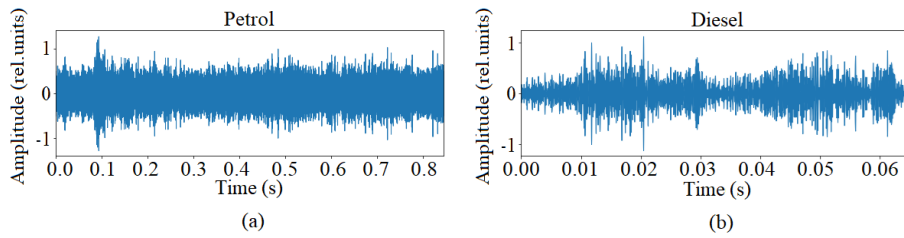


Fig. 13 Audio signal of (a) petrol and (b) diesel engine at idle extracted from the recordings with an early change of operation mode

As the mapping of audio signals into an adequate image format [21, 22], such as spectrogram-like images, is increasingly used in modern signal processing and deep learning, the obtained audio signals are also presented in the form of spectrograms, see Figs. 14, 15 and 16. There are some properties present in the spectrograms of both engine types (petrol and diesel), such as stronger components at low and mid frequencies than at high frequencies as well as rather steady-state behavior along the time axis. However, these images contain also certain differences between the sounds of petrol and diesel engines, such as more uniform energy distribution along frequency axis for the petrol engine and more prominent particular frequency components for the diesel engine. More detailed analysis of the recorded audio signals and their representations in different domains, as well as correlation between the signals and vehicle types by fuel will be done in the very next phase of the research.

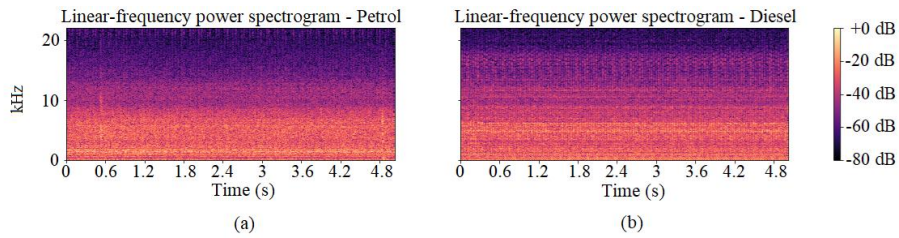


Fig. 14 Spectrogram of (a) petrol and (b) diesel engine audio signal at idle, without changing the mode and without applying the idling mode extraction

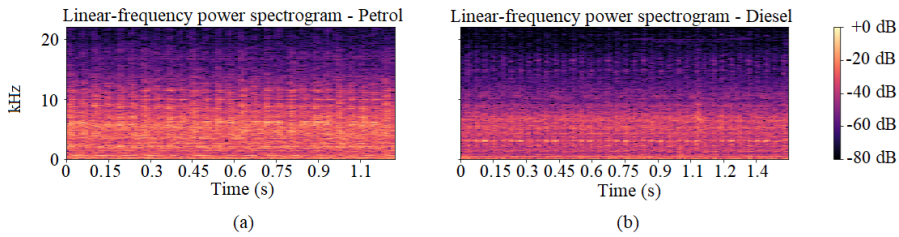


Fig. 15 Spectrogram of (a) petrol and (b) diesel engine audio signal at idle with a late change of operation mode

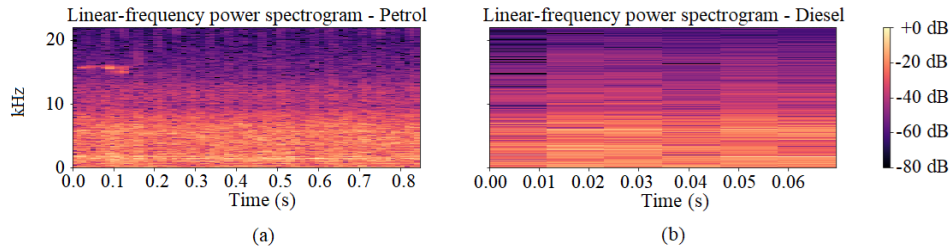


Fig. 16 Spectrogram of (a) petrol and (b) diesel engine audio signal at idle with an early change of operation mode

4. CONCLUSIONS

Considering the number of recordings containing exclusively the idling mode of the vehicles in reference to the number of sampled vehicles, it can be seen that the developed acquisition system has collected at least one such recording for each vehicle. Also, the system has not recorded a single blank audio file, and it is rather robust to false triggering. In addition, the selected amount of memory proved to be sufficient, and the most critical part of the system, the battery power, gave very satisfactory results in terms of system autonomy. Since 250 vehicles in total passed behind the microphone and sensors placed on the ground without any consequences for functionality, the condition of robustness has been satisfied, and also the ability of unattended use has been proven.

The developed additional processing of recorded signals for extracting exclusively the engine idle mode along the entire audio recording has enabled to create a dataset of audio samples containing only this target mode of operation. The acquisition system has proven to be efficient for recording the sound of a passenger vehicle at idle regardless of the type of fuel. The number of audio recordings can also be affected by the approach applied for detecting the presence of a vehicle using ultrasound sensors. This results in a larger or smaller number of recordings having higher or lower redundancy between the recordings, respectively.

By using the developed acquisition system, a dataset has been created consisting of 352 audio recordings for 250 vehicles containing the sound of engines in the idling mode of operation. This acquisition system can find its application in different use-cases including control of car entrance in restricted areas of smart-cities, prevention of misfueling at gas stations, optimization of road usage or noise prevention based on engine fuel type. In such cases, this proof-of-concept system could be implemented as an embedded system on a dedicated single platform.

Depending on a particular application and its requirements, the acquisition system might be modified to become even less demanding. Thus, taking into account relatively high sound pressure levels at the microphone (above 74 dB) and proximity of the source, the condenser AKG C562CM microphone might be replaced by an electro-dynamic microphone not requiring phantom power. Since it is expected that dynamic range of the acquired signals will not be that large, the bit depth might be smaller than 16 bits used here. In addition, after developing an adequate classifier and considering the useful frequency range, it would be worthwhile to explore an option of reducing the sampling frequency.

The generated dataset of audio samples will play an important role in future work for developing a system for automatic recognition of the type of engine based on the used fuel. This system will be designed by applying an adequate approach of deep or machine learning for classification and employing the created dataset for model training and testing. Based on the samples from the generated dataset, it can be concluded that spectrograms of engines that use petrol and diesel at idle seem to be different, forming a strong ground-base for achieving high accuracy in engine type classification.

Acknowledgment: *This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, contract no. 451-03-68/2022-14/200102.*

REFERENCES

- [1] S. Das, A. Dey, A. Pal and N. Roy, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", *Int. J. Comput. Appl.*, vol. 115, no. 9, pp. 31-41, April 2015.
- [2] P. Dhanalakshmi, S. Palanivel and V. Ramalingam, "Classification of Audio Signals Using SVM and RBFNN", *Expert Syst. Appl.*, vol. 36, no. 3, part 2, pp. 6069-6075, 2009.
- [3] P. Dhanalakshmi, S. Palanivel and V. Ramalingam, "Classification of Audio Signals Using AANN and GMM", *Appl. Soft. Comput.*, vol. 11, no. 1, pp. 716-723, 2011.
- [4] H. Ponce, P. Ponce and A. Molina, "Adaptive Noise Filtering Based on Artificial Hydrocarbon Networks: An Application to Audio Signals", *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6512-6523, 2014.
- [5] Z. Liu, J. Huang, Y. Wang and T. Chen, "Audio feature extraction and analysis for scene classification", In *Proceedings of First Signal Processing Society Workshop on Multimedia Signal Processing*, Princeton, NJ, USA, 23-25 June 1997, pp. 343-348.

- [6] T. Birtchnell, "Listening Without Ears: Artificial Intelligence in Audio Mastering", *Big Data & Society*, vol. 5, no. 2, July 2018.
- [7] G. P. Chossière, R. Malina, F. Allroggen, S. D. Eastham, R. L. Speth and S. R. H. Barrett, "Country- and Manufacturer-Level Attribution of Air Quality Impacts due to Excess NO_x Emissions from Diesel Passenger Vehicles in Europe", *Atmospheric Environ.*, vol. 189, pp. 89-97, Sept. 2018.
- [8] M. Milivojčević, F. Pantelić, D. Ćirić, "Pozicioniranje mikrofona prilikom snimanja audio karakteristika motora putničkih vozila" (Microphone positioning when recording audio characteristics of passenger car engines) In Proceedings of 63rd National Conference on Electrical, Electronic and Computing Engineering ETRAN, Srebno Jezero, Serbia: 3-6 June 2019, pp. 58-62 (in Serbian).
- [9] M. Milivojčević, F. Pantelić and D. Ćirić, "Comparison of frequency characteristics of sound generated by internal combustion engines depending on fuel", In Proceedings of 26th Noise and Vibration, Niš, Serbia: 6-7 December 2018, pp. 115-120.
- [10] N. Evans, *Automated Vehicle Detection and Classification using Acoustic and Seismic Signals*. Ph.D. Thesis, University of York, 2010.
- [11] H. Frederick, A. Winda and M. Iwan Solihin, "Automatic petrol and diesel engine sound identification based on machine learning approaches", In Proceedings of the International Conference on Automotive, Manufacturing, and Mechanical Engineering. Bali, Indonesia: 26-28 September 2018, published at E3S Web of Conferences, vol. 130, article no. 01011.
- [12] A. D. Mayvana, S. A. Beheshtib and M. H. Masoom, "Classification of Vehicles Based on Audio Signals using Quadratic Discriminant Analysis and High Energy Feature Vectors", *Int. J. Soft Comput.*, vol. 6, no. 1, pp. 53-64, Feb. 2015.
- [13] A. Wiczorkowska, E. Kubera, T. Słowik and K. Skrzypiec, "Spectral Features for Audio Based Vehicle and Engine Classification", *J. Intell. Inf. Sys.*, vol. 50, pp. 265-290, 2018.
- [14] E. Alexandre, L. Cuadra, S. Salcedo-Sanz, A. Pastor-Sánchez and C. Casanova-Mateo, "Hybridizing Extreme Learning Machines and Genetic Algorithms to Select Acoustic Features in Vehicle Classification Applications", *Neurocomput.*, vol. 152, pp. 58-68, March 2015.
- [15] S. D. Badiger and M. UttaraKumari, "Vehicle Classification Using Machine Learning Algorithms Based on the Vehicular Acoustic Signature", *Sci. Tech. Dev.*, vol. 8, no. 11, pp. 369-374, Nov. 2019.
- [16] Ultrasonic Waterproof Range Finder datasheet. Available at: <https://www.jahankitshop.com/getattach.aspx?id=4635&Type=Product>.
- [17] A. Pajankar, *Kickstart to Arduino Nano*. Susteren, The Netherlands: Elektor International Media, 2022.
- [18] B. R. Kent, *Science and Computing with Raspberry Pi*. San Rafael, USA: Morgan & Claypool Publishers, 2018.
- [19] C562 CM specifications. Available at: <https://www.akg.com/Microphones/Boundary%20Layer%20Microphones/C562CM.html>.
- [20] Digital high definition microphone interface specifications. Available at: <https://www.ikmultimedia.com/products/irigprehd/>.
- [21] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, A. Baird and B. Schuller, "Snore sound classification using image-based deep spectrum features", In Proceedings of Interspeech 2017, Stockholm, Sweden, August 20-24, 2017, pp. 3512-3516.
- [22] D. Ćirić, Z. Perić, J. Nikolić, N. Vučić, "Audio signal mapping into spectrogram-based images for deep learning applications", In Proceedings of 20th International Symposium Infoteh-Jahorina (INFOTEH), East Sarajevo, Bosnia and Herzegovina: March 17-19, 2021, pp. 1-6.